Quality Assessment Report:

GTG version 1.0 Proposed Threshold Changes for Aviation Digital Data Service Displays

Michael P. Kay[1,2], Judy K. Henderson[1], and Jennifer L. Mahoney[1]

[1]NOAA-Research/Earth System Research Laboratory, Boulder, CO
[2]Cooperative Institute for Research in Environmental Sciences/University of Colorado, Boulder, CO

Quality Assessment Product Development Team

14 August 2006

1. INTRODUCTION

The Federal Aviation Administration Aviation Weather Research Program's (AWRP) Turbulence Product Development Team (TPDT) has proposed altering the set of thresholds used to categorize turbulence from the Graphical Turbulence Guidance (GTG) on the Aviation Digital Data Service (ADDS) displays. GTG, as run operationally at the Aviation Weather Center, was calibrated using input data from the Rapid Update Cycle (RUC) model, which ran with a horizontal resolution of 20 km. On June 28th, 2005, the RUC model was upgraded to run at a higher horizontal resolution of 13 km. GTG forecasts, and the corresponding displays on ADDS have not been altered to account for the changes in calibration associated with the increased horizontal resolution of the RUC model.

This report summarizes the differences in performance of the operational GTG using a proposed new set of forecast thresholds for categorizing turbulence intensity from the performance of GTG found using the existing operational thresholds. GTG forecasts are designed and calibrated to predict moderate or greater (MOG) turbulence events. While MOG turbulence is the focus of the predictions, all observed turbulence intensities from voice pilot reports (PIREPs) are used in the design of the forecasts (Sharman et al. 2006). GTG data have values ranging from zero to unity; intermediate values within this range are associated with predicted categories of turbulence intensity such as light, moderate, and severe.

GTG forecasts are typically provided to users through ADDS in the form of graphical images that display the spatial location of predicted turbulence either at a given flight level or through an integrated layer of the troposphere such as 20,000 – 45,000 ft (Fig. 1). These images do not display contours of raw GTG output, but instead depict the following categories of expected turbulence intensity: none, light, moderate, severe, and extreme; where none represents all intensities less than light (including smooth/none).

This report is organized in the following way. Section 2 provides an overview of the data and approach for evaluating the thresholds. Section 3 presents the results of the comparison. The conclusions and discussions are presented in Section 4.

2. DATA AND METHODOLOGY

This section explains the data and methods used to assess the quality of GTG at categorical forecasts of turbulence intensity. Data for the period 1 November 2005 through 31 January 2006 were used for the evaluation. Valid times between 1500 UTC and 0000 UTC were used in order to maximize the number of PIREPs available. Table 1 shows the set of issue times and lead times available for GTG during the evaluation period.

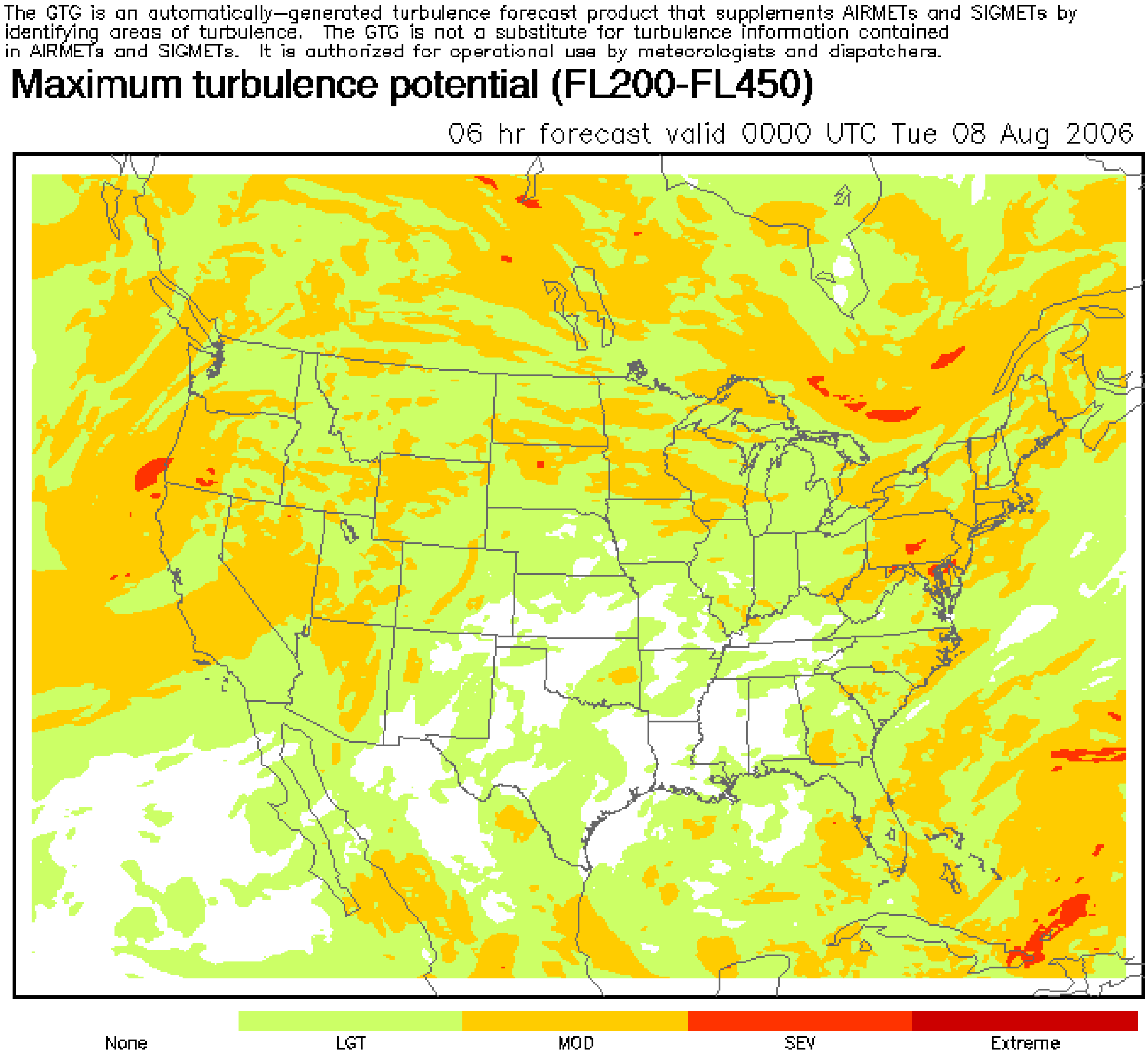

Fig. 1: ADDS display of turbulence location and intensity for the 6-h forecast from GTG valid 0000 UTC 08 Aug. 2006 for the 20,000 – 45,000 ft layer.

Table 1: The set of issue times and lead times for GTG used in this report.

Issue Time (UTC)	Lead Time (h)
1200	3, 6, 9, 12
1500	3, 6, 9
2100	3

The basic approach to creating forecast and observation pairs is identical to past studies of turbulence performed by the Quality Assessment Product Development Team (e.g., Takacs et al. 2004). Forecasts are interpolated in three dimensions to PIREP locations for all PIREPs that occur within a ± 60 min. window of the forecast valid time. The forecast- and observation intensities are then transformed into categories using the mapping described in Table 2. This mapping was applied to both the current as well as proposed thresholds of GTG. The corresponding categories can then be aggregated to provide the joint distribution of forecasts and observations that describe the non-time dependent forecast quality information of GTG using the different thresholds for categorization. Due to the lack of data for extreme turbulence, any forecasts or observations for extreme events are included in the Severe category.

Table 2: Mapping of forecast and observed values to categories used in this report. Values represent lower bounds of the ranges of data associated with each category.

Category	GTG (Current)	GTG (Proposed)	PIREP Intensities
None	0.0	0.0	Smooth/None, Smooth to occasional light
Light	0.125	0.3	Light to occasional moderate
Moderate	0.375	0.475	Moderate, Moderate to occasional severe
Severe	0.625	0.75	Severe, Severe to occasional extreme, Extreme

Data were gathered over the Continental U.S. in addition to nearby coastal waters, corresponding to the Aviation Weather Center forecast regions. Within this region, the forecast volume of turbulence greater than or equal to each of the thresholds from Table 2 was computed. Percent volume was computed by dividing the total volume forecast by the total volume possible within the airspace.

3. RESULTS

The joint distribution of forecast and observations provides the relevant information necessary to discern the performance of GTG at categorical predictions of turbulence. The goal of the forecasting system (i.e., GTG) is to perfectly match the forecast category with the correct observed category. When viewed in a tabular manner, all elements along the diagonal would have non-zero values and all off-diagonal elements would be zero. Two additional derived distributions can be created from the raw distributions: the conditional probability of an observation given a forecast (denoted p(x|f)), and the conditional probability of a forecast given an observation (denoted p(f|x)). The former distribution, p(x|f), will serve as the focus for the analysis since the primary question to be answered is whether changing the GTG thresholds for each category leads to better predictions.

GTG performance using the existing thresholds is shown in Table 3. It is important to reiterate that the results presented here are only valid for the locations where PIREPs existed. No definitive statements can be made about the true nature of GTG performance throughout the airspace as there are no systematic observations of turbulence that encompass all spatial and temporal scales. In Table 3, results should be viewed on a row-by-row basis owing to the conditioning process. For example, when Light turbulence was forecast by GTG, 69.5% of the time an intensity of None (PIREPs of None/Smooth and Smooth-to-Occasional-Light) was reported in the PIREP, 9.3% of the time the PIREP reported Light turbulence, 20.6% the PIREP reported Moderate turbulence, and 0.6% of the time the PIREP indicated Severe turbulence. Therefore, GTG performed well for forecasts of None. For other categories of turbulence, such as Moderate and Severe, GTG had difficulty clearly discriminating between these observation intensities. For instance, the Severe category was not forecast well by GTG; only 1.7% of the forecasts and observations properly agreed and more than 25% of Severe forecasts were actually associated with observations of no turbulence.

Table 3: Conditional probability of an observation given a forecast, p(x|f), using the existing thresholds for defining the categories.

		Observed					
		None	Light	Moderate	Severe	N	p(f)
Forecast	None	**0.846**	0.051	0.101	0.002	19064	0.339
	Light	0.695	**0.093**	0.206	0.006	27041	0.480
	Moderate	0.501	0.133	**0.355**	0.011	9490	0.169
	Severe	0.251	0.164	0.568	**0.017**	720	0.013

In spite of the difficulty with clearly categorizing the PIREP intensities, the proposed new set of thresholds appear to slightly improve the performance of GTG as indicated by the results in Table 4. The forecasts of None and Moderate best correlate with the respective PIREP categories of None and Moderate, as indicated by p(x|f) values of 77.6% and 45.0%, respectively. Due to the small number of PIREPs available for evaluation of the Severe threshold, little can be said about the performance of the Severe forecast category using the new threshold value of 0.75. Moderate forecasts were accompanied by Moderate PIREPs 45.0% of the time but were also associated with None PIREPs 38.8% of the time. Forecasts of Light were most often verified with PIREPs of None (59.3%) but were also associated with a large percentage of Moderate PIREPs (28.1%). The changes in the ability of GTG to predict the correct category of observation owing to the threshold change are shown in Table 5.

Table 4: Conditional probability of an observation given a forecast, p(x|f), using the proposed thresholds for defining the categories.

		Observed					
		None	Light	Moderate	Severe	N	p(f)
Forecast	None	**0.776**	0.071	0.149	0.004	40800	0.724
	Light	0.593	**0.118**	0.281	0.008	10626	0.189
	Moderate	0.388	0.150	**0.450**	0.013	4887	0.087
	Severe	0.500	0.000	0.500	**0.000**	2	3.5e-5

The changes in the ability of GTG to predict the correct category of PIREP intensity as the result of the threshold change are shown in Table 5. None forecasts from GTG are now accompanied by None PIREPs in 77.6% of the situations, which is a decrease from 84.6% for the current threshold. This decrease is due to more forecast values now falling into the None category as a result of the increase in the Light threshold from 0.125 to 0.3. The None category remains the best predicted category from GTG. However, improved forecasts for the Light and Moderate categories are also evident. For instance, Moderate forecasts show the most improvement with GTG correctly predicting Moderate PIREP observations rising from 0.355 to 0.450. Neither of the two PIREPs associated with Severe forecasts were rated of Severe intensity. Therefore, p(x|f=Severe) was 0.00 and the difference negative. Again, little weight should be placed on this result because of the insufficient sample size (2).

Table 5: Difference in p(x|f) for GTG using the current- and proposed thresholds for defining turbulence categories.

Category	p(x|f) (Current)	p(x|f) (Proposed)	Difference (Proposed-Current)
None	0.846	0.776	-0.07
Light	0.093	0.118	0.025
Moderate	0.355	0.450	0.095
Severe	0.017	0.000	-0.017

Despite the improvements associated with the change in thresholds, GTG still does not correctly predict the observation intensity in many situations (cf. Table 4). There are several reasons that may explain this behavior. First, the PIREP dataset is dominated by observations of intensities None and Moderate. Results presented above indicate that GTG does indeed perform best for the None and Moderate categories in terms of correctly predicting these observation categories. Calibration of GTG is tied to the PIREP dataset and therefore may also be linked to that dataset's apparent bias towards the None and Moderate intensities. Second, GTG was explicitly designed for the prediction of moderate-or-greater (MOG) turbulence only. Discrimination of MOG turbulence from less-than-MOG turbulence is clearly a difficult problem. Further insight into GTG behavior can be found by looking at the distributions of forecast values conditioned upon the observed PIREP intensity, or category (Fig. 2). For the Light, Moderate, and Severe categories, the median forecast values are all very similar near 0.3. The shapes of the distributions vary slightly, but are quite similar with long tails on the right-hand side indicating the presence of larger forecast values associated with that particular observed category. Observations of category None show a clear distinction from the other categories; the median forecast value is approximately 0.15. In this view, GTG does a reasonably good job of discriminating between light-or-greater turbulence and None (none/smooth-to-occasional-light).

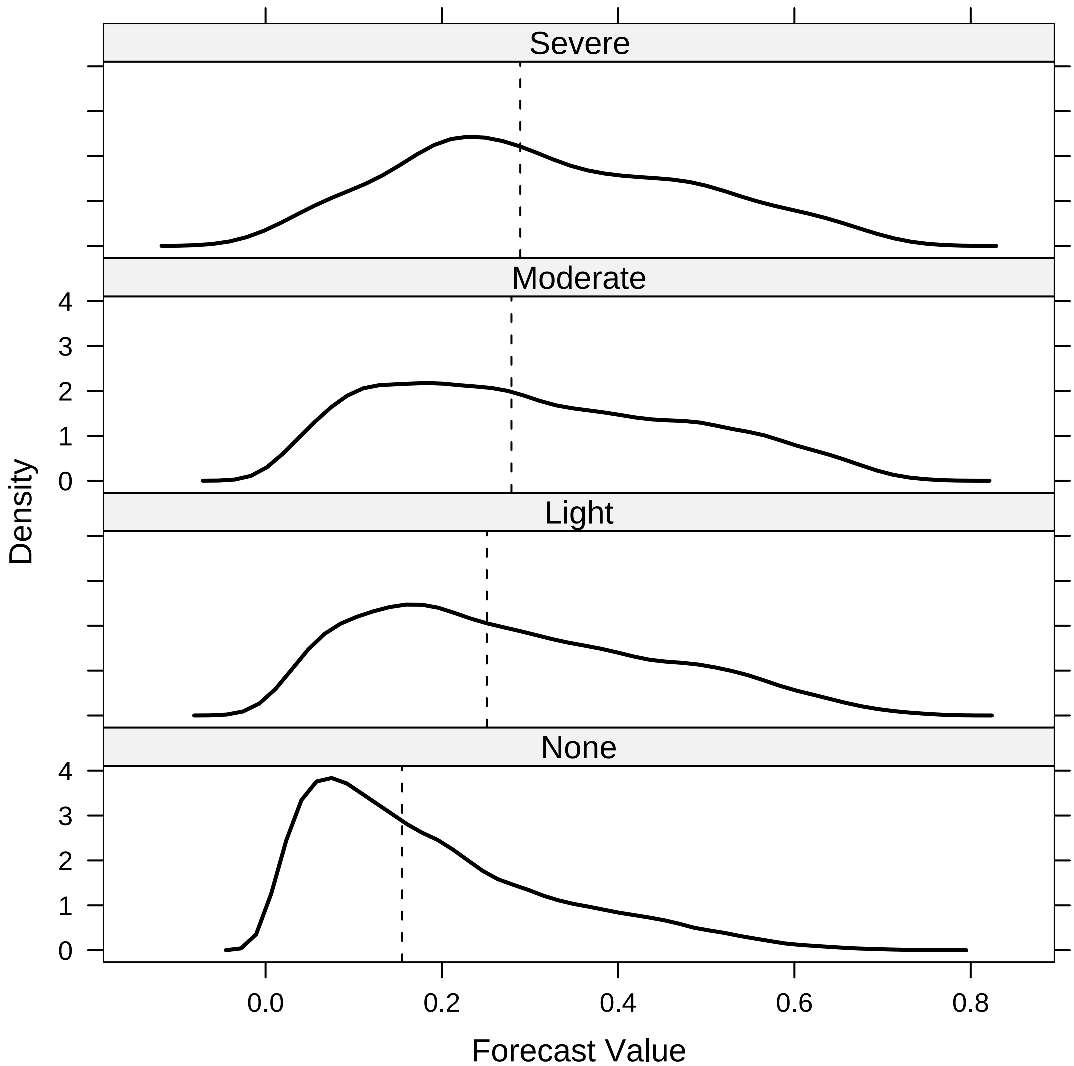

Fig. 2: Probability density functions of forecast values associated with each category of observations. Vertical dashed lines indicate the median values of each distribution.

Another aspect of performance to consider is the effect of the change in thresholds on the volume of turbulent airspace predicted (Table 6). Increasing the lowest threshold, from 0.125 to 0.3 for Light-or-greater turbulence decreases the volume of impacted airspace by 71.1%, from 46.3% to 13.4%. MOG volumes decreased by over 60%, while volumes for Severe-or-greater events decreased by 99%. While the true amount of turbulence in the free atmosphere is unknown, the decreased volumes using the new thresholds are clearly an improvement in forecast performance and are in line with previous estimates of turbulence in the atmosphere (Vinnichenko et al. 1980).

Table 6: Median Percent Volume values for GTG using the current and proposed thresholds for categorization of turbulence intensities.

Category	Median Percent Volume (Current)	Median Percent Volume (Proposed)	Percent Change
Light-or-greater	46.3	13.4	-71.1
Moderate-or-greater	7.7	3.1	-60.4
Severe-or-greater	0.26	0.002	-99.3

4. CONCLUSIONS

GTG categorical forecast performance was analyzed with respect to proposed changes in thresholds to define the categories for display on ADDS. The results indicated that the proposed new thresholds for the intensity categories appear to improve the quality of the GTG forecasts by reducing the impacted airspace and by improving the categorization of turbulence intensity. More specifically, GTG performed moderately well with respect to predicting the correct intensity of turbulence, especially for the Moderate and None categories. The volume of airspace affected by GTG forecasts decreased dramatically with decreases of more than 60% for all categories, when the proposed thresholds were applied.

5. REFERENCES

Sharman, R., C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Weather and Forecasting*, **21**, 268-287.

Takacs, A., L. Holland, M. Chapman, B.G. Brown, J.L. Mahoney, and C. Fischer, 2004: Graphical Turbulence Guidance 2 (GTG2): Quality Assessment Report. Submitted to Aviation Weather Technology Transfer (AWTT) Technical Review Panel.

Vinnichenko, N.K., N.Z. Pinus, S.M. Shmeter, and G.N. Shur, 1980: *Turbulence in the Free Atmosphere*. Plenum, 310 pp.